ILIAD As An Expert Consultant to Teach Differential Diagnosis

Homer R. Warner, MD, PhD, Peter Haug, MD, Omar Bouhaddou, PhD, Michael Lincoln, MD, Homer Warner, Jr, Dean Sorenson., PhD, John W. Williamson, MD, Chinli Fan, MS

Department of Medical Informatics, University of Utah School of Medicine
Salt Lake City, Utah 84132

ABSTRACT

A program (ILIAD) is described which uses knowledge frames representing diseases encountered in internal medicine to teach medical students about differential diagnosis. ILIAD consultant utilizes a number of sophisticated inferencing mechanisms to mimic the strategy of a medical expert in working up a patient. Its knowledge is represented in both Bayesean and Boolean frames which permit use of sensitivities and specificities to describe the relationship of a disease to its manifestations and provide a basis for explaining its conclusions. In addition to differential diagnosis, ILIAD provides advice regarding the most appropriate information to seek at each stage of the workup. ILIAD's knowledge base is also used to simulate patient cases and evaluate the problem solving performance of medical apprentices. It is presently being used by third year medical clerks on the wards of three hospitals and beta tested at eight additional sites.

INTRODUCTION

Expert systems such as HELP [1], QMR [2], MEDAS [3] and DXplain [4] use a knowledge frame to represent the relationship of each disease to all that disease's manifestations. The length of such a list, however, makes it of limited value as a means of helping a student learn to recognize the disease process in a patient. In addition, such a list of manifestations usually involves dependencies that violate the model, whether the model is based on statistical parameters or empirical weights.

The most successful statistical models applied to diagnosis in medicine have been based on Bayes theorem [5]. A frame-based version of the Bayes model which formulates the problem in terms of deciding whether a patient does or does not have each disease [6] allows for more than one disease to be present in a patient. This is used by ILIAD. However, the sequential Bayesean decision model also requires that the disease manifestations be independent of each other in patients with the disease. To avoid the limitations imposed by this constraint, we have defined clusters of manifestations which are not independent and are usually caused by some common underlying process and often given a name by clinicians (i.e., pleuritic chest pain', 'typical angina', 'signs of systemic infection'). The logic used to express relationships among items in these clusters must not only deal with the dependency among the manifestations but also represent the expert clinicians logic in a form that can be learned by a medical student. We have chosen to use Boolean expressions for this purpose.

GOAL OF PROJECT

The goal of the project is to provide the medical student during his/her third year clerkship on internal medicine with immediate expert consultation and advice that will facilitate learning the problem-solving skills required of a good physician. To accomplish this we have built a microcomputer-based system (ILIAD) that can mimic an expert diagnostician. ILIAD is able to recognize each of the diseases we expect a medical student to become familiar with. It allows free-text entry of observations made by the student during his/her workup of the patient, can provide consultation to the student at any stage of the process regarding the differential diagnosis, and can advise the student on the most appropriate observation to make next. In addition, ILIAD utilizes its knowledge base to generate "synthetic" patient cases and evaluates the student's workup strategy by comparing it to its own under the same circumstances.

GENERAL FEATURES OF THE SYSTEM

Since ILIAD was designed primarily as a tool to supplement the student's learning experience while working on the wards with real patients, design of the system led us to think of ways to present medical knowledge to students that; 1) would be easy to understand, 2) could be remembered, 3) would facilitate communication to others in the profession, and 4) made use of the students knowledge of the underlying pathophysiology of disease.

To achieve these ends, the program (ILIAD) includes the following functional features:

1. ILIAD can draw conclusions from incomplete information.
2. Less specific information can be used until specific data becomes available.
3. It uses probability-based decision logic with formal provision for insuring independence
4. An explanation can be presented at any stage of a patient work-up showing how the program arrived at its working hypotheses or why it is asking for certain information.
5. The user may view directly in an easy-to-understand hypertext format the logic in any frame in the knowledge base at any point in the problem-solving process.
6. Upon request, ILIAD will provide the user with advice as to which information would be most cost-effective to seek next.
7. It provides a user with the facility to change an item of information (and thereby undo all the effects of that item) at any point.
8. A case may be interrupted at any point and continued when new information is available.
9. Notes may be saved with the case to facilitate knowledge-base refinement to overcome problems encountered with the system.
10. The hierarchial structure of the system dictionary is used by ILIAD to avoid requesting data that can be inferred.
11. ILIAD supports an open-ended, random sequence of data input.
12. Program performance improves as statistical estimates in the knowledge base are updated from a patient database which contains the program's cumulative "experience".

BASIC COMPONENTS OF THE SYSTEM

Inference engine: The inference engine of ILIAD is independent of the subject area being addressed in a particular application. It controls the communication with the user, evokes the appropriate knowledge frames to handle the information entered, requests information needed to test hypotheses that have been evoked, and explains its conclusions to the user on request.

User interface: Through the use of pull-down menus and windows, the user may control the operation of the program to meet his/her needs. With the 'add data' option, the student is presented with a window in which to type one or more words or partial words describing a finding. Several findings may be entered in the same window separated by a semicolon. ILIAD uses a stripping algorithm to try to identify a suffix in each word and then from the stem term looks in its dictionary for a concept matching the terms in each entered string. If more than five hits are found, the user is so informed and may see them all or may add additional constraining words to his/her query. Otherwise, the hits are presented for the user to make one or more selection sor enter a numerical value where appropriate.

Data driver: With each frame added to the knowledge base, a pointer is created from each dictionary item used by that frame to the appropriate slot in the frame. Thus, during the workup of a patient, each new piece of information acquired about the patient will automatically evoke the logic that uses that information. If the value of the frame itself is changed, any other frames that use information from this frame will be processed as well. Thus, all ramifications of each new observation will be pursued.

Best information: The scoring algorithm used by ILIAD for ordering best information questions is:

The score of an item of information = Inf * Prob / C

where:

Inf, the information gain for this frame expected from this item, = MAX {abs(Ln(Sens/(1-spec))), abs(Ln((1-sens)/spec))},

Prob = the apriori probability of the frame being true (i.e. before getting the item of information), and

C = the dollar cost to acquire this item of information.

KNOWLEDGE BASE

The knowledge-base component of ILIAD is the subject-specific knowledge and consists of the dictionary of terms used to describe the subject and a set of frames (tables) written in a loosely-structured text frame format.

Hierarchical dictionary:: Much of the knowledge is contained in the hierarchical structure of the dictionary. Each item in the dictionary has a text string describing it and a code which uniquely identifies its position in the hierarchy. This permits many of the inferencing functions described below.

Inference: The program is capable of making several types of inferences based on the hierarchical structure of the data dictionary. For example, if a user states that a patient has bloody sputum, the program will infer that the patient has a cough and will not ask the user for that information. On the other hand, if the user tells the program that the patient does not have a cough, then the program will not ask if the patient has bloody sputum.

Probability of the symptom (Ps), cost: Each item in the system dictionary has a number stored with it that represents the expected frequency (Ps) with which that feature or attribute occurs in the subject population from which our patients come (the combined population of three hospitals; University Hospital, LDS Hospital and VA Medical Center). Ps can be estimated from:

$$Ps = \text{sum} \{ Pd_i * Ps/d_i \}$$
over all diseases

where

Pd_i is the apriori probability (prevalence) of the i^{th} disease,

and Ps/d_i is the probability of a patient with the i^{th} disease having manifestation 's'.

Each dictionary item also has a cost stored with it which will be used by the inference engine in choosing which item to pursue. This cost is the actual cost to the patient (or provider) in most cases. History items are set to $1 and physical exam items to $2.

CLUSTERS

In this paper we describe an approach to modeling a disease by grouping its manifestations into subsets (syndromes or clusters) which represent entities well known to clinicians. This approach is considered to closely emulate the logical analysis used by domain experts in making medical decisions in practice. In many cases such a subset of manifestations represents a pathophysiological process that may be shared by several diseases. The actual criteria used in designing cluster frames are obtained from the domain expert. There may be disagreements even among experts as to these criteria, but personal biases can usually be minimized by discussion among several experts. Starting with these original expert opinions, the system can later be improved based on experience using data from real patients. This approach does not involve arbitrary weights that cannot be confirmed by experience with the system. If the expert cannot define a rule, either the cluster may not be real or some other model for the frame should be sought.

A cluster frame is designed as a decision module built around a Boolean relationship among its findings, i.e., any one or some combination of findings in the list may be sufficient for the frame to come true. For example, we might say that a patient has 'Signs of Systemic Infection' if he has two or more of the following findings: fever, chills, or night sweats. This frame represents a nonspecific collection of findings commonly associated with bacterial infections, and a variety of disease conditions, in which the items are not independent, and in which it is not necessary to grade the findings in terms of severity. This is the natural way for a clinician to describe such an entity; that is, as the minimum combination of findings that would justify attaching the name of the cluster to a given patient. In this example, the cluster also represents a pathophysiological process, i.e., a number of different types of bacteria localized in the gastrointestinal tract or pulmonary system release substances which are absorbed into the blood stream and consequently have effects on white blood cells, smooth muscle and the brain which eventually result in one or more of the findings listed in the cluster frame.

```
TITLE  Acute MI
TYPE  probability

PREVALENCE (apriori) = 0.04

FINDINGS                                             disease   no disease
a. @7.142.101 Risk of Coronary Artery Disease
       low risk      0.0 to 0.25                       .05        .45
       medium risk   0.26 to 0.45                      .25        .35
       high risk        > 0.45                         .70        .20
a. @7.142.111 Infarction chest pain                    .85        .01
     else
   @7.142.114 Unstable angina pain                     .40        .02
     else
   @7.142.112 Typical angina pain                      .10        .02
     else
   chest pain                                          .90        .19
b. @7.142.120 Left-sided heart failure                 .20        .05
c.  @7.149.113 Autonomic reaction to stress            .42        .10
d. ECG: acute myocardial infarction                    .88        .01
e. CPK/MB elevation (>7% of total) within 24
    hours after pain onset                             .70        .02
     else
   @7.149.112 Signs of tissue necrosis                 .40        [illegible]
```

Figure 1. Example of probability based decision frame for the diagnosis of acute myocardial infarction

The Acute Myocardial Infarction (AMI) frame (Figure 1) exemplifies some additional important features in our system. The @ sign designates cluster frames. In other words, 85% of patients with AMI, also had findings sufficient to satisfy the criteria for infarction chest pain (Figure 2); only 1% of patients without AMI satisfied this criteria. The Infarction chest pain cluster could, in turn, include other clusters in its list of findings, in a nested fashion.

TITLE : Infarction chest pain
TYPE interpretation

FINDINGS

a. chest pain, substernal
b. chest pain, lasting > 15 minutes
c. chest pain, heavy/pressure/aching
d. chest pain, relief with rest/NTG
e. chest pain, pleuritic or positional
f. chest wall tenderness

True if (a or c) and b and not d, e or f

Figure 2. Example cluster frame called by the frame acute myocardial infarction

Many diseases are defined independently of their clinical findings from anatomical, etiological, or biochemical parameters and by their known natural course and even their response to treatment. The association of a disease with its pathophysiological and clinical manifestations may vary from patient to patient and even in a given patient from time to time. Thus, a statistical model seems the most appropriate way to represent such a diagnostic decision process.

In our system, decision or disease frames are written in probabilistic terms whenever possible. This makes the analysis more quantitative and objective and facilitates the ranking and comparison of multiple hypotheses. It is necessary, however, before proceeding to design probability frames, to delineate the setting in which the frames will be used. In other words, one must define the disease and non disease populations before obtaining probabilities. These definitions must then always be kept in mind when obtaining probabilities from the expert, the patient database, or the literature. For the purposes of teaching medical students diagnoses, we assume the total population to be all of the patients in our hospital (or those admitted to the medical wards) and the non-diseased population, all those patients without the disease being modeled in a given frame.

A typical probability disease frame for acute myocardial infarction would contain the following slots (Figure 1): 1) an apriori probability, 2) a list of findings by which the disease may be recognized, and 3) the frequency of these findings in patients with and without this disease. For the case of acute myocardial infarction, we found an apriori probability of as many as three in 100 in our hospital. After delineating the list of findings thought to be necessary and sufficient to diagnose acute myocardial infarction, the sensitivities are determined or estimated as the fraction of patients who end up with that diagnosis who actually have that finding at the time they are being diagnosed. The corresponding values for the nondiseased population are the fraction of patients with some other diagnoses who have that finding. The ratio of these two numbers for each finding represents the importance of that item in the frame. The larger the magnitude of that ratio, either positive or negative, relative to the other items in the frame, the more important the original item is in making the final decision. On the other hand, if the ratio is close to 1, the item is of little value in making the decision, even if it is a finding commonly associated with the condition being considered. The importance of this conclusion can't be overemphasized because of the tendency for inexperienced clinicians to diagnose a condition based on the presence of a symptom with a very high sensitivity even though the symptom may also be present in numerous other conditions.

Risk factors: Another use for a cluster is to delineate a list of predisposing factors. For example, we might say that a patient has had a significant silica exposure if he worked in one of several occupations for a period of at least 10 years or alternatively, if he worked as a sandblaster or tunnel builder for only one year. Another example would be a list of drugs which might predispose someone to an adverse reaction to a drug under consideration for treatment. This type of cluster frame allows the knowledge engineer to separate a long list of possible related factors from the main decision frame even though those items may be independent. This then simplifies the decision frame and makes it more understandable to a student since the cluster 'explains' why the findings in the cluster are important to the diagnosis.

Else structure: There can be alternate rules for a frame, depending upon which combination of findings are present at the time of execution of the frame. The else designator is another method we use to circumvent the problem of nonindependence. For example, if a patient with AMI has infarction chest pain, it is also quite possible that he will have unstable angina pain as well. In cases where two or more elsed statements are satisfied, only the one with the most importance is used by the computer program in the calculation to generate the differential diagnostic list.

ILIAD SIMULATION MODE

ILIAD is an expert system which can be used in a simulation mode to create hypothetical cases whose purpose is to teach and test medical decision-making skills. During this mode, ILIAD can evaluate the diagnostic capabilities of the student by operating in its consultation mode on the same patient observations seen by the student and scores the deviation of the student's workup strategy to its own.

A Patient Case Simulator (PCS) module generates a set of patient observations from a disease frame in ILIAD's knowledge base using a random number generator; the observed frequencies associated to each manifestation of the illness and the general incidence of the other terms in the system dictionary (Ps). For example, in the disease frame for AMI (Figure 1) the observed frequency for "Left-sided Heart Failure" is 0.20. A status present is assigned to the finding if the random number drawn from a uniform distribution on the interval (0, 1) is below 0.20 and absent if otherwise. When a finding is described as a multi-bin variable (e.g., "Risk of Coronary Artery Disease" in AMI, Figure 1), the random number is generated from a distribution adjusted to the finding's empirical distribution. In the case of an "else structure", the PCS observes the hierarchy between the elements involved as it does with all dictionary items. In the AMI frame item b, the PCS will first attribute a status to "Chest Pain" then, if "Chest pain" is present, the PCS will assign a status to "Typical angina pain" and so on, from general to specific information. The PCS handles cluster frames in a recursive way. Given the sensitivity of the cluster in the disease frame, a status (present or absent) is assigned to the cluster then, a status is assigned to each element of the cluster based on its role in confirming or denying the cluster. Empirical scores are derived from the cluster Boolean logic expression to describe the contribution of each element. For example, if the Boolean expression to confirm the cluster reads: "If finding a AND finding b" then, the empirical scores of "finding a" and "finding b" are equal and equal to 0.50.

We believe that ILIAD simulation mode can provide a potentially useful educational tool. Its implementation has required no additional restructuring of the knowledge base which could be a major undertaking [7] and only minimal programming efforts and its use helps validate the structure and adequacy of the knowledge base. ILIAD simulation mode is now being used to evaluate medical students diagnostic skills prior to their exposure to ILIAD consultation mode.

EVENT DRIVEN MACINTOSH APPROACH

Implementation: The program is implemented on the Macintosh personal computer. The computer program is written in 'C' and is easily stored on a single floppy disk. The program itself is menu driven and easy to use. A tutorial and documentation are provided with the program which is now being commercialized by the a.l.p. Systems (Salt Lake City, Utah) and is presently being tested in eight locations.

Menus: Although the program allows for random input of data through the 'add data' mode, it also determines, by itself, what the next best item to be input should be in the 'most useful information' mode. This is achieved by ranking each item in each frame under consideration according to an algorithm which takes into account the probability of the frame in which the item resides, the information about the frame provided by the item, and the cost of that information.

```
<< 0.638 Infective Endocarditis >>
Could explain the following:
   ??--  Infective embolic phenomena
     Yes -- Clubbing of fingers
  No -- Have you had a heart valve replaced?
  Yes -- Have you had a fever with this illness?
  Yes -- Signs of systemic infection
       Yes -- Malaise?
       Yes -- Fatigue (weak and tired)?
       Yes -- Have you been sweating at night?
       Yes -- Have you had a fever with this illness?
       Yes -- Have you had chills with this illness?
  Yes -- Have you had a fever for at least two weeks?
<< 0.988 =.== aortic insufficiency >>
  Yes -- Have you ever had rheumatic fever?
   Yes -- Diastolic blowing or high-pitched murmur
            at the left sternal border?
   No -- Apical diastolic rumble?
    ?? -- Left-sided heart failure
       Yes -- Low cardiac output
          Yes -- Fatigue or malaise?
       ?? -- Pulmonary venous congestion
          Yes -- Rales
Age -- 53
Sex -- male
```

Figure 3. Example explanation presented to the student in the "Why" mode during runtime operation of ILIAD

If the user desires an explanation of the logic of a decision frame (similar to the one in Figure 3), he/she may access the frame for any given disease by a menu option or may select any disease in the differential diagnostic list and inquire as to which of the findings entered contributed to its position in the list as well as the underlying hierarchical relationship between the decision frames, clusters, and findings involved. In another window, a list of the findings entered by the user is present on the screen at all times. If the user selects one of the findings in this list, an explanation of this finding is presented in the form of an ordered set of clusters and diseases which most likely account for this abnormality

Windows: A sequential Bayesian calculation is used to generate the differential diagnostic list. Note that, with this format, the observations (patient findings) can be input in any order. In each case a single Bayesian calculation is done and the apriori probability is modified accordingly. The result of this calculation may then change the ranking of this disease on the differential diagnostic list. If at some later time, another item in an else list with more importance than one previously used in a calculation becomes available, the program undoes the calculation performed with the original manifestation and recalculates the probability with the more specific information. This facility also makes possible a 'change information' feature in ILIAD. That is, if the user at some time during the investigation wants to go back and see what would have been the outcome if a previously entered finding had been true rather than false or vice versa, the program will immediately make the appropriate calculations and re-rank the diseases in the differential diagnostic list.

Another useful feature offered by the program is to save all the information entered at any point so that a user can recall the case at some later time and continue with the workup. Saving the cases provides a record of each student's experience with the system. The user may enter comments to be saved with the case at any point in the process.

Setting for implementation on wards: Three hospitals provide facilities for third year clerkship experience for students at the University of Utah. These are the University, LDS, and VA hospitals. Macintosh computers, each with hard disk and a printer, have been installed on each medical ward of all three hospitals. Each case entered by each student is saved in a database and provides a record for future reference and for correlation with the students performance on test material at the end of the clerkship. The student may access the machine during rounds and may prepare printouts of any pertinent material. Residents and attending physicians may call for the student's computer record of the workup and question the student and his 'consultant'.

An experiment supported by a grant (#5 R01 LM04604) from the National Library of Medicine is underway to evaluate the effect this tool will have on student learning. This will be the subject of a future SCAMC paper.

REFERENCES

[1] Warner HR, Olmstead CM, Rutherford BD. HELP--a program for medical decision-making. Computers and Biomedical Research 1972; 5:65-74.

[2] Miller RA, Maserie FE, & Myers JD. Quick medical reference (QMR) for diagnostic assistance. MD Computing 3:34-48 (1986).

[3] Ben-Bassat M. Pattern-based interactive diagnosis of multiple disorders: the MEDAS system. IEEE Press 1980; Transactions on pattern analysis and machine intelligence PAMI-2(2):148.

[4] Barnett GO, Cimino JJ, Hupp JA, & Hoffer EP. DXplain: Experience with knowledge acquisition and program evaluation. 11th Annual SCAMC, Washington D.C. (1987), pp150-154.

[5] Warner HR, Toronto AF, Veasy LG. and Stephenson R. A mathematical approach to medical diagnosis--application to congenital heart disease. JAMA 177:177 (1961).

[6] Warner HR. Computer-assisted medical decision-making Academic Press, New York, NY 1978.

[7] Parker RC, Miller RA. Using causal knowledge to create simulated patient cases: the CPCS project as an extension of INTERNIST-I. SCAMC 1987; pp473-480.